AD-A066 661 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
GENERALIZED APPROACH FOR PREDICTING A DICHOTOMOUS CRITERION.(U)
FEB 79 J R DEMPSEY, W S SELLMAN, J C FAST
UNCLASSIFIED AFHRL-TR-78-84 NL

| OF |
AD
A066661

END
DATE
FILMED
5-79
DDC

AFHRL-TR-78-84

LEVEL II

2

AIR FORCE

HUMAN RESOURCES

AD A066661

DDC FILE COPY

GENERALIZED APPROACH FOR PREDICTING A DICHOTOMOUS CRITERION

By

Jack R. Dempsey, Capt, USAF
Wayne S. Sellman, Major, USAF

AIR FORCE MILITARY PERSONNEL CENTER
Randolph Air Force Base, Texas 78148

Jonathan C. Fast, Capt, USAF

OCCUPATION AND MANPOWER RESEARCH DIVISION
Brooks Air Force Base, Texas 78235

February 1979
Interim Report for Period October 1977 – March 1978

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This interim report was submitted by Occupation and Manpower Research Division, under project 2077, with HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235. Capt Jonathan C. Fast (ORS) was the Principal Investigator for the Laboratory.

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

RAYMOND E. CHRISTAL, Technical Director
Occupation and Manpower Research Division

RONALD W. TERRY, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFHRL-TR-78-84 (14)
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): GENERALIZED APPROACH FOR PREDICTING A DICHOTOMOUS CRITERION. (6)
5. TYPE OF REPORT & PERIOD COVERED: Interim rept. October 1977 – March 1978, (9)
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Jack R. Dempsey, Wayne S. Sellman, Jonathan C. Fast (10)
8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas 78235
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, (16) 2077 0407, (17) 04
11. CONTROLLING OFFICE NAME AND ADDRESS: HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235
12. REPORT DATE: February 1979 (11)
13. NUMBER OF PAGES: 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 20p.
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
attrition
dichotomous dependent variable
maximum likelihood estimation
personnel modeling
prediction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report refines and improves upon a conceptual model and a mathematical procedure, based upon a blend of Likelihood Function Estimation (LIFE) and utility theory. Empirical studies conducted by the Air Force Military Personnel Center in 1975 and 1976 have shown that the LIFE procedure can be very useful in the prediction and study of dichotomous behavior, e.g., predicting attrition/success of a trainee in an Air Force Training program. The technique has been previously used to study attrition from the United States Air Force Academy and Basic Military training. By generalizing the LIFE technique it can be applied to the situation of studying and predicting any dichotomous dependent variable. Currently, the Air Force Human Resources Laboratory is continuing studies of the procedure to compare its usefulness to other mathematical methods.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified 404 415

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

79 03 30 081

PREFACE

The contents of this technical report reflect the results of research done primarily at the Air Force Military Personnel Center within the office of the Assistant for Personnel Plans, Programs, and Analysis, during 1975 and 1976. The efforts of the co-authors, Jack R. Dempsey, Wayne S. Sellman, and Jonathan C. Fast, were previously reported in two published technical memorandums (see Dempsey & Fast, 1976; Dempsey & Fast, 1977). The purposes of this technical report are to refine the previous mathematical presentations, to make the research available to potential users on a wider basis, and to serve as a basis for research currently being undertaken at the Air Force Human Resources Laboratory.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

GENERALIZED APPROACH FOR PREDICTING A DICHOTOMOUS CRITERION

I. INTRODUCTION

Many occasions arise in research where the dependent criterion is of a dichotomous or binary nature (e.g., a pass/fail criterion, where an individual either succeeds or fails). Traditionally, researchers have attacked this problem using ordinary least squares (OLS) regression. Many statisticians and econometricians have critized this application of OLS as being unappropriate and theoretically unsound (see, for example, Nerlove & Press, 1973). This paper presents an alternative approach which uses a mathematical model that is theoretically better founded than OLS in the case of the dichotomous criterion. The model described in this report uses the Likelihood Function Estimation (LIFE) technique, which maximizes this function to develop predictions of the dependent dichotomous criterion. In section II, the mathematical description of the LIFE model is developed, and in section III different methods for interpreting and applying the model are presented. The previous research done by the authors, using personnel data to describe whether a person succeeds or fails in a training program, is contained in the appendices to this report. Appendix A summarizes research which used Air Force Academy cadets as subjects and which was previously reported in Dempsey and Fast (1976). Appendix B describes research which used first-term airmen accessions to the Air Force and which was previously presented in a paper at the OSD/ONR [Office of the Secretary of Defense/Office of Naval Research] Conference on First Term Attrition, 4–7 April 1977 (Dempsey, Fast, & Sellman, 1977).

II. THE LIFE MODEL

Let Y be a dichotomous random variable defined to be 1 if an event occurs and 0 otherwise. Let X be an m × n matrix of m explanatory variables of Y which may be dichotomous, polytomous, or continuous. Let β be m × 1 vector of coefficients such that $(X'\beta)_i$ specifies a linear function of X, for each observation (i = 1, . . . ,n). Finally let ξ denote an n × 1 vector of random disturbances distributed N(0,1). By hypothesis, Y is related to X such that:

$$Y_i = 1 : \text{when } (X'\beta)_i + \xi_i > U_i \text{ (event occurs)}$$

$$Y_i = 0 : \text{when } (X'\beta)_i + \xi_i \leq U_i \text{ (event does not occur)}$$

where U_i represents an n × 1 vector of random variables that can be interpreted in different ways. For the purposes of this development, there will be no interpretation; this is discussed further in section III. The random variables U_i are assumed to be distributed $N(0,\sigma^2)$.

Let P_i represent the probability of an event E occuring such that:

$$P_i = \text{Prob } [(X'\beta)_i = \xi_i > U_i] \tag{1}$$

which can be expressed further by (2):

$$P_i = \int_{-\infty}^{+\infty} \int_{-\infty}^{\xi_i + (X'\beta)_i} f(\xi_i, U_i)\, dU_i\, d\xi_i \tag{2}$$

where $f(\xi_i, U_i)$ is the joint density function of ξ_i and U_i. Since there is a systematic component, $(X'\beta)_i$, and a random component, $U_i - \xi_i$, this can be reduced to a more manageable form by making the substitution $Z_i = U_i - \xi_i$. The new component will be distributed $N(\mu', \sigma'^2)$,

where:

$$\mu' = E(\mu) - E(\xi) \qquad (3)$$
$$= 0 - 0 = 0$$

and

$$\sigma'^2 = VAR(\mu) + VAR(\xi) \qquad (4)$$
$$= \sigma^2 + 1$$

Equation (2) reduces to:

$$P_i = \int_{-\infty}^{(X'\beta)_i} f(Z')dZ' \qquad (5)$$

The standardized random variable can then be defined as:

$$Z = \frac{Z' - \mu'}{\sigma'} = \frac{Z'}{\sigma'}$$

$$dZ = \frac{1}{\sigma'} dZ'$$

Then (5) reduces to:

$$P_i = \int_{-\infty}^{\frac{(X'\beta)_i}{\sigma'}} f(Z)dZ \qquad (6)$$

Since $f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}Z^2\right)}$

$$P_i = \int_{-\infty}^{\frac{(X'\beta)_i}{\sigma'}} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}Z^2\right)} \qquad (7)$$

P_i then is the value of the normal distribution CDF evaluated at the point $\frac{(X'\beta)_i}{\sigma'}$. This can be written as $F\left(\frac{(X'\beta)_i}{\sigma'}\right)$

The following substitutions are made for notational convenience:

$$\text{Let } J_i = \frac{(X'\beta)_i}{\sigma'} \qquad i - 1, \ldots, n$$

$$\text{Let } \alpha_k = \frac{\beta_k}{\sigma'} \qquad k = 0, \ldots, m$$

$$\text{Let } \alpha_{m+1} = \frac{1}{\sigma'}$$

$$\text{Let } I_i = X'\beta_i + \xi_i$$

The Maximum Likelihood Solution

Since the probability of each occurrence P_i is specified for each $i = 1, \ldots, n$, the likelihood function can be formed, and the estimate of the β_i can be found which maximizes the likelihood function for this sample. Let the sample of n observations be ordered, where the first r observations equal zero and the remaining $n - r$ observations equal 1. Without loss of generality, the likelihood of the sample is given by:

$$L = \prod_{i=1}^{r} [1 - F(J_i)] \cdot \prod_{i=r+1}^{n} F(J_i)$$

The natural logarithm of this function is given by:

$$\ln L = \sum_{i=1}^{r} \ln[1 - F(J_i)] + \sum_{i=r+1}^{n} \ln F(J_i)$$

Let X_0 be exactly 1 for all i. Then setting the partial derviatives of lnL, with respect of α_k, equal to 0 yields the following system of m+1 equations:

$$\frac{\partial \ln L}{\partial \alpha_k} = \sum_{i=1}^{r} \frac{-f(J_i)}{[1 - F(J_i)]} X_{ki} + \sum_{i=r+1}^{n} \frac{f(J_i)}{F(J_i)} X_{ki} = 0$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\partial \ln L}{\partial \alpha_{m+1}} = \sum_{i=1}^{r} \frac{-f(J_i)}{[1 - F(J_i)]} (X'\beta)_i + \sum_{i=r+1}^{n} \frac{f(J_i)}{F(J_i)} (X'\beta)_i = 0$$

These equations are non-linear but can be solved using any one of several iterative techniques. The solution yields a set of b_i, estimates of the maximum likelihood coefficients, $\frac{\beta_k}{\sigma'}$, and s, an estimate of σ'. These coefficients are used to form:

$$\hat{I}_i = S(X'b)_i$$

and

$$\hat{P}_i = F[(X'b)_i]$$

an estimate of P_i for each observation.

III. INTERPRETATION AND DEVELOPMENT

At this point, the LIFE model has developed a probability of occurrence for the dichotomous criterion studied. For many purposes, this will be sufficient and can serve a very useful purpose. For example, in the case where the criterion was the attrition from or success in an Air Force training program, the probability developed can be interpreted to be the probability of attrition from the training and could be used in a selection method for rank ordering individuals. However, in many applications the researcher wishes to predict the outcome (0 or 1) of the criterion. In this case, the $\hat{P}_i$ must be used to produce a predicted dichotomous outcome for each observation. There are two methods for developing this outcome: empirical observation and fiducial inference.

Empirical Observation

Using this method, the original sample is reordered by sorting on $\hat{P}_i$, the estimate of P_i for each observation. A cut score, C_0, is then developed for the sample using some optimality criterion developed by the researcher. If $\hat{P}_i > C_0$, then the event is said to occur, i.e., $\hat{Y} = 1$. If $\hat{P}_i \leq C_0$, then event is predicted not to occur, i.e., $\hat{Y} = 0$. The optimality criterion could be based on the cut score which achieves the most correct classifications ($Y_i = 0$ and $\hat{Y}_i = 0$ or $Y_i = 1$ and $\hat{Y}_i = 1$). Another criterion which could be used would be a trade-off between a low false positive rate ($\hat{Y}_i = 1$ and $Y_i = 0$) and a high correct classification of failures ($\hat{Y}_i = 1$ and $Y_i = 1$). The optimality criterion, however, should be chosen to meet the needs of the manager and the program for which the prediction system is being developed.

Fiducial Inference

Another method for developing the prediction system would be to interpret the random variable, U_i, in the special case where the observations are actually occurrences as a result of human behavior. In this case, where an individual is exercising his or her choice mechanism to decide on which alternative to take, U_i can be interpreted to be the utility function described in the classical Marshallian framework (Marshal, 1961). "The attractiveness of a trade depends not on its money earnings, but its net advantages." Initially, the individual surveys the available alternatives and weighs the advantages and disadvantages of each. Naturally the individual selects the one with the highest net advantages. Consider for example the recurring decision facing the Air Force Academy cadet. Assume the cadet makes an implicit dollar valuation for a current career choice and a similar valuation for an alternative choice, given the cadet's view of each. So long as the subjective dollar valuation of the current career choice (Academy utility) is greater than the subjective dollar valuation of the alternative career choice (alternative utility), the cadet remains at the Academy. As long as the net difference in utilities is positive, the choice is made to remain in the Academy; where the net difference is negative, the alternative occupation is chosen.

This utility theory framework can be used to infer within some fiducial limit what the outcome will be for each individual. In estimating β, it has been assumed that the X vector is a vector of fixed variables. This constraint may be relaxed as long as it is assumed that X is uncorrelated with β, ξ, and U. By relaxing this assumption, it may be said that the utilities among individuals for the alternative choices are distributed as independent bivariate normal random variables. Then the probability density function of I and U is given as:

$$f(U_i,I_i) = f_1(U_i)f_2(I_i)$$

Let $W_i = I_i - U_i$. W_i represents the difference between the respective utilities and will determine which alternative the individual chooses. The interest then is in finding the distribution of this difference function W_i. Using the convolution formula, this density function can be found.

$$g(W_i,U_i) = f(U_i,W_i+U_i)\frac{dU_i}{dW_i}$$

Integrating U_i from $-\infty$ to $+\infty$, $g(W_i)$ is given by:

$$g(W_i) = f(W_i + U_i, U_i)\, dU_i$$

This can be simplified to:

$$g(W_i) = \int_{-\infty}^{+\infty} f_1(W_i + U_i) \cdot f_2(U_i)\, dU_i$$

where

$$f_1 = f_2 = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\left[1/2\left(\frac{\mu}{\sigma}\right)^2\right]}$$

)

Thus, the density of W_i is:

$$g(W_i) = \frac{1}{\sigma^* \sqrt{2\pi}}\, e^{-1/2} \left(\frac{\omega_i - \mu^*}{\sigma^*}\right)^2$$

where:

$$E(\mu^*) = \int_{-\infty}^{+\infty} I_i f(I_i) - \int_{-\infty}^{\infty} U_i f(U_i)$$

$$= \beta_0$$

and:

$$\sigma^* = \sigma' + \left(\sum_{i=1}^{m} \sigma_i^2\right)^{1/2}$$

$$\sigma_i = \text{Std Dev of } X_i$$

Considering that W_i represents the difference between the respective utilities, when the difference equals zero, the individual is said to be indifferent between the two alternative choices. Thus $g(0)$ is the mean point of difference for all individuals and is given by $F(\beta_0)$, which can be estimated by $F(sb_o)$.

To use this estimate, three uncertainties must first be accounted for: (a) uncertainty in the mean point of indifference, (b) uncertainty in the estimators, and (c) uncertainty in the random disturbances. First the upper confidence bound on the estimator b_0 is constructed:

$$b_0^* = sb_0 + sz_\alpha \left[\mathrm{VAR}(b_0)\right]^{1/2}$$

Then, the lower confidence bound on the estimator $\hat{I}_i$ is constructed, given X.

$$\hat{I}_i^* = \hat{I}_i - Z_\alpha \left[\sum_{i=0}^{m} VAR(b_j)X_j + 1 \right]^{1/2}$$

The prediction is then made under the following regime:

If $f(\hat{I}_i^*) > F(b_0^*)$, the event is predicted to occur, i.e., $\hat{Y} = 1$.

If $f(\hat{I}_i^*) \leq F(b_0^*)$, the event is predicted not to occur, i.e., $\hat{Y} = 0$.

IV. CONCLUSION

The mathematical method and the conceptual model presented in this report offer a unique blend of utility theory and likelihood estimation techniques. This combined model represents a useful alternative for the study and prediction of dichotomous behavior of individuals in Air Force Training programs. In addition, the mathematical technique can be generalized for the prediction and description of any dichotomous or binary dependent variable.

REFERENCES

Dempsey, J.R., & Fast, J.C. *Predicting attrition: An empirical study at the United States Air Force Academy.* AFMPC-TM, AD-A024 816. Randolph AFB, TX: Air Force Military Personnel Center, March 1976.

Dempsey, J.R., Fast, J.C., & Sellman, W.S. *A method to simultaneously reduce involuntary discharges and increase the available manpower pool.* In H.W. Sinaiko (Ed.), *First term enlisted attrition* (Vol. I). Washington, D.C.: Manpower Research and Advisory Services, Smithsonian Institute, June 1977.

LaChar, D., Sparks, J.C., & Larsen, R.N. Psychometric prediction of behavioral criteria of adaption for USAF basic trainees. *Journal of Community Psychology*, 1974, **2**(3), 268–277.

Marshal, A. *Principles of economics* (8th ed.). London: MacMillian, 1961.

Nerlove, M., & Press, S.J. *Univariate and multivariate log-linear and logistic models.* Santa Monica, CA: Rand Corporation, 1973.

BIBLIOGRAPHY

Berkson, J. Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 1944, **39**, 357–365.

Berkson, J. A statistically precise and relatively simple method of estimating the bio assay with quantal response, based on logistic function. *Journal of the American Statistical Association*, 1953, **48**, 565–599.

Finney, D.J. *Probit analysis.* Cambridge, England: Cambridge University Press, 1947.

Goldberger, A.S. *Econometric theory* (2nd ed.). New York: John Wiley & Sons, 1964.

Sonquist, J.A., & Morgan, J.N. *The detection of interaction effects.* Ann Arbor, MI: University of Michigan, Institute for Social Research, 1969.

Their, H. *Principles of econometrics.* New York: John Wiley & Sons, 1969.

Tobin, J. Estimation of relationships for limited dependent variables. *Econometrics*, 1958, **26**, 24–36.

Wonnacott, R.J., & Wonnacott, T.H. *Econometrics*. New York: John Wiley & Sons, 1970.

APPENDIX A: RESULTS OF STUDIES AT THE UNITED STATES AIR FORCE ACADEMY

I. THE UNITED STATES AIR FORCE ACADEMY INITIAL STUDY

This section describes an initial test conducted at the United States Air Force Academy and designed to evaluate the conceptual approach and estimation procedure used in this report for potential application to other Air Force programs. The Air Force Academy was selected to test the methodology because of the extensive data maintained on each candidate/appointee/cadet.

Background

Historically, the Air Force Academy has experienced a cadet attrition rate which has ranged between 28 and 46 percent. An estimated two-thirds of these cadets possess a significant motivational component whereby the separation action is initiated by the individual. The remaining attrition can be roughly classified as either academic or miscellaneous. Academic attrition generally results from formal board action after the cadet has failed to meet the minimum academic standards for retention, while miscellaneous separations include such reasons as hardship, medical, and accidental death. Upon separation, each cadet has his record annotated with a two digit code which (cross-referenced to a master list) best describes his reason for leaving. Since the conceptual model precludes involuntary action on the part of the cadet, this initial test was designed to predict only motivational (voluntary) attrition.

Data

The data used included information from four major sources—The Air Force Academy General Information Questionnaire (GIQ), the Survey of High School Activities (HSA), the Strong Vocational Interest Blank (SVIB), and other data relating prior academic achievement.

General Information Questionnaire (GIQ): The GIQ is a questionnaire designed to provide both personal background data and information about factors that influenced the candidate to apply to the Academy. The GIQ is mailed to the candidate for completion and is returned to the Academy prior to arrival of the candidate.

Survey of High School Activities (HSA): The purpose of the HSA is to provide information about each appointee's participation in extracurricular activities while in high school; included are varsity sports and fraternal and elective organizations. The survey is completed by each cadet within 2 weeks of arrival at the Academy.

Strong Vocational Interest Blank (SVIB): The SVIB is a 399 item self-report inventory that assesses a cadet's interest in various occupational and general interest areas. Eight-four scales can be constructed using responses to items that have been previously identified as being related to specific occupations.

Prior Academic Achievement: A transcript of each candidate's high school academic record is transmitted to the Academy and includes course grades and class standing. In addition performance on the College Entrance Examination Boards (CEEB), Scholastic Aptitude Test (SAT), or American College Test (ACT) are sent to the Academy. These scores are weighted to develop several indices which are used in the selection process: prior academic record (PAR), scientific index, and non-scientific index. Other indices are generated which incorporate additional non-academic information: athletic index, non-athletic index, leadership composite, weighted composite, and academic composite.

Test Methodology

Certain data elements were extracted from the four primary data sources which were then used to construct a record on each cadet. Each record was annotated with the cadet's status as of 1 June 1975 (0 if still enrolled, 1 and discharge code if not enrolled). Any record which was missing one or more of the principal variables was eliminated from the sample.

The test was conducted using the classes of 1976 and 1977. A prediction equation and critical limit (prediction system) were estimated for the class of 1976 using the estimation procedure discussed in this report. This prediction system was then applied to the class of 1977 for cross-validation. Table A1 shows the sample sizes for the two classes.

Table A1. **Sample Sizes for Initial Test**

Category	Year of Class 1976 N	Year of Class 1977 N
Cadets Still Enrolled	916	937
Motivational Attritions	237	246
Total in Sample	1,153	1,183

Results

The LIFE procedure correctly classified 32.1 percent of the actual attritions and 94.2 percent of the actual successes (Table A2). Figure A1 shows that over 59 percent of the predicted attrition group did, in fact, leave the Academy within their first 2 years while only 15.8 percent of the predicted success group separated. All of these separations were classified by the Academy as possessing a significant motivational component.

Table A2. **Prediction Results Class of 1977**

Category	Predicted Attritions	Predicted Successes	Total	Percent Correct
Actual Attritions	79	167	246	32.1
Actual Successes	55	882	937	94.2
Total	134	1,049		
Percent Correct	59.0	84.2		

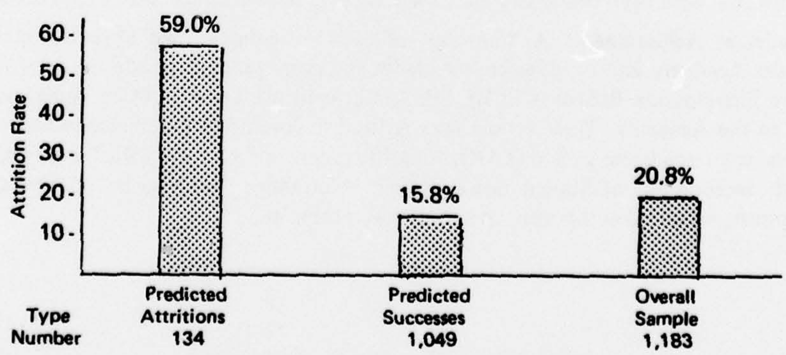

Figure A1. **Attrition rates class of 1977.**

II. THE UNITED STATES AIR FORCE ACADEMY EMPIRICAL TEST

This section describes a test to evaluate the conceptual approach and estimation procedure for possible application to other Air Force programs.

Background

Based on the results of the initial test described in the previous section the feasibility of the approach had been demonstrated. The empirical test described herein was designed to demonstrate that the methodology could, in fact, predict attrition a priori on a by-name basis. It was important to evaluate the procedure in a simulated operational environment which would require a 2 year lag in the prediction system. For these reason, the empirical test was conducted using the class of 1977 to estimate the prediction equation and critical limit and using the class of 1979 as the demonstration class.

Data

The empirical test utilized the same data and format collected for the class of 1977 in the initial test. Identical data were collected on the class of 1979 and a similar record constructed for each cadet. However there was one difference in the method of construction. Any cadet record missing one or more of the principal variables was discarded from the sample in the initial test. Because the purpose of the empirical test was to simulate an operational environment in which all candidates would receive a prediction, any record missing a principle variable was given the mean value of that data element. This resulted in a 99.8 percent sample of the entering class of 1979 (Table A3).

Table A3. **Sample Sizes for the Empirical Test**

Category	Year of Class 1977 N	1979 N
Cadets Still Enrolled	937	1,257[a]
Motivational Attrition	247	178
Total in Sample	1,183	1,460[b]

[a]At completion of test.

[b]Total in 1979—there were also 25 attritions for other reasons.

Test Methodology

A prediction system was estimated using the class of 1977 and was then applied to the members of the Class of 1979 within 3 weeks after their arrival. The duration of the empirical test was approximately 6 months which allowed sufficient time to adequately assess the performance of the procedure. The test was terminated on 12 December 1975.

Results

The procedure was able to correctly classify 36.0 percent of the motivational attritions and 91.3 percent of the actual successes (Table A4). Over 37 percent of the predicted attritions had separated by the end of their first semester (Figure A2). Thirteen additional predicted attritions separated shortly after their return from Christmas leave; seven of these were motivational.

Table A4. **Prediction Results Class of 1979**
(Including Only Motivation Attritions)

Category	Predicted Attritions	Predicted Successes	Total	Percent Correct
Actual Attritions	64	114	178	36.0
Actual Successes	110	1,147	1,257	91.3
Total	174	1,261		
Percent Correct	3.70	91.0		

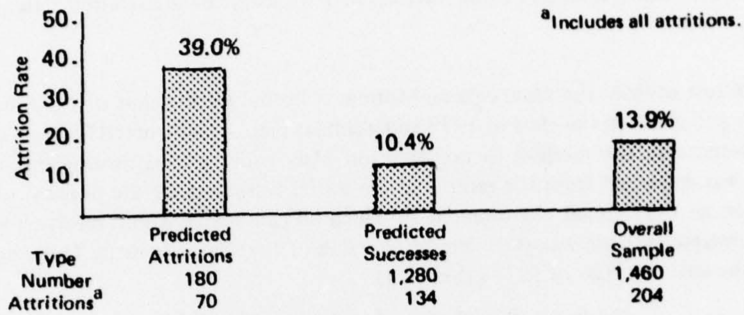

Figure A2. **Attrition rates class of 1979.**

APPENDIX B: PREDICTING ATTRITION AMONG NON-PRIOR SERVICE FIRST TERM ACCESSION

Using the LIFE Model to Derive a More Precise Enlistment Standard

The uncertainty in current Service enlistment standards and the favorable results obtained at the United States Air Force Academy provided the impetus to investigate whether the LIFE model could be used to derive a more efficient enlistment standard for the Air Force.

The Sample

The sample population consisted of 14,923 Air Force accessions who entered the Service between June and August 1972.

Procedure

To obtain discharge data, the data file maintained by the Computational Sciences Division, Air Force Human Resources Laboratory was matched with airman tape files maintained by the Air Force Military Personnel Center. A total of 607 cases in the original population did not match the official data files, and eliminating these reduced the sample population to 14,316. The loss of these cases is not thought to materially bias the analysis presented.

Discharge status was determined by official loss code which identified all personnel who had been separated from the Service during the first term of enlistment. Loss codes indicating a voluntary/normal loss were grouped together as were loss codes indicating a discharge of an involuntary nature. Based on the specific loss code each individual was assigned to one of three mutually exclusive groups (Table B1).

Table B1. **Categories of Sample**

	Group	Sample Size
I	Active Duty	10,002
II	Voluntary Loss	669
III	Involuntary Loss	3,645
	Total	14,316

Since most voluntary/normal losses do not result from marginal performance or adverse behavior, voluntary/normal losses were removed from the sample in order to isolate the effect of enlistment criteria on involuntary losses exclusively. The removal of this group further reduced the sample population to 13,647.

Because the LIFE algorithm restricts the number of observations to 3,000 or less, a computational sample of 2,642 was randomly selected from the sample population (Table B2).

Table B2. **Categories of Random Sample**

	Group	Sample Size
I	Active Duty	1,992
II	Involuntary Loss	650
	Total	2,642

Model Specification

After performing a series of preliminary analyses using Automatic Interaction Detection (AID), the following data model was specified.

Independent Variable	Transformation
X_1 = age at enlistment (years)	0 if $X_1 \geq 19$, 1 otherwise
X_2 = education level (years)	0 if $X_2 \geq 12$, 1 otherwise
X_3 = Administrative composite plus electrical composite	Standardized score
X_4 = *Military Service Inventory* (*MSI*)[1]	Standardized score
X_5 = Number of dependents in household	0 if $X_5 \leq 2$, 1 otherwise
X_6 = Armed Forces qualifying test	Standardized score

With the model specified, a utility function and indifference point were estimated for the computational sample using the LIFE model (Table B3).

Table B3. **Estimated Coefficients and T-Value**

Variable		Coefficient	T-Value
b_1	Age	.125707	1.87
b_2	Education Level	.355775	2.51
b_3	Administrative & EI	−.037114	1.69
b_4	MSI	.343853	2.43
b_5	Number Dependents	.283619	1.76
b_6	AFQT	.034158	1.71
U_i (Indifference Point) = .52		α = −.650289	

Comparative Analysis

Once the utility function and indifference point were estimated using the LIFE method, the coefficients and indifference point were used to weight the appropriate selection data and establish a cutting score respectively. The original sample of 13,647 CY 72 accessions was then rescreened using this standard. But to make the results more meaningful with respect to impacts on recruiting and attrition, the sample population was rescreened using the current Air Force enlistment standards and several other hypothetical, but traditionally oriented, enlistment standards (Figure B1).

Discussion of Results

According to the analysis presented in Table B4, the LIFE standard had the highest pass-rate, lowest loss-rate, and did not adversely affect the quality of enlistees. In fact, 57% of the individuals who would have been denied enlistment if a LIFE standard had been used in CY 72 were involuntarily separated prior to completion of their first term of enlistment. This means that out of every 100 individuals that would have been denied enlistment under the LIFE standard in CY 72, only 43 would have succeeded. This compares to 62 potentially successful applicants turned away under current Air Force enlistment standards.

---

[1] The Military Service Inventory (MSI) is a 50 question self-report inventory developed by the authors. The development of the MSI was spinoff of a previous study conducted by LaChar, Sparks, and Larsen, 1974, who developed a psychometric instrument called the History Opinion Inventory (HOI) for the purpose of identifying airmen who would be unable to adapt to a military environment. The 100 questions contained in the HOI were revalidated against a criterion of involuntary attrition and restructured into a 50 question format.

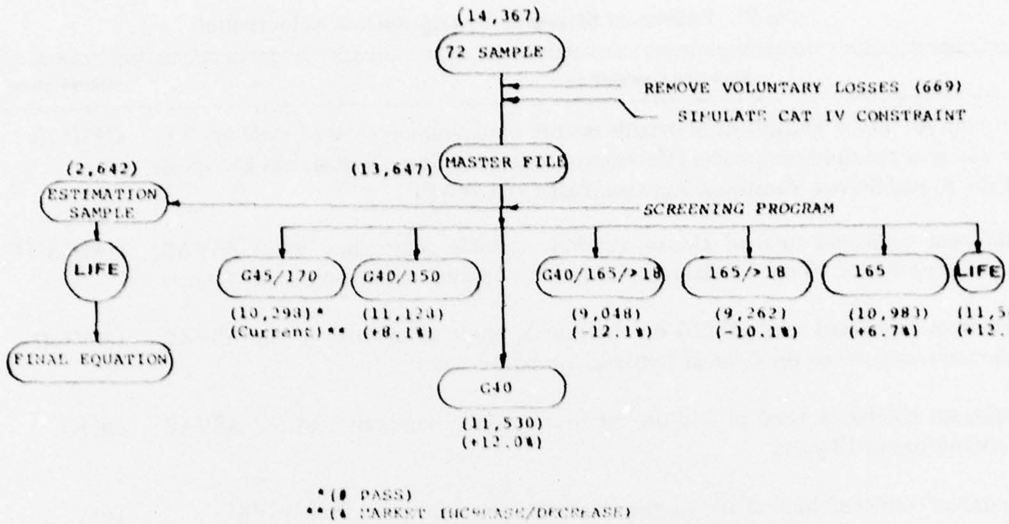

Figure B1. **Methodology of Analysis.**

Table B4. **Comparison Chart**

# Standard[a]	Pass Rate	Loss Rate	Quality Indicators										Force Characteristics	
				ASVAB Average				Average AFQT	Mental Category					
			<HS	M	A	G	E		I	II	III	IV	Minority	Average Age
1 G45/170	75%	23%	6%	64	63	68	68	66	6%	48%	45%	1%	9%	18.8
2 G40/165/>18	66	22	4	64	62	68	68	66	7	46	46	1	10	19.1
3 G40/150	82	24	6	64	61	66	67	65	6	46	47	1	11	18.8
4 165/>18	68	22	4	64	62	67	67	65	6	45	47	2	10	19.1
5 165	80	23	6	63	61	66	67	65	6	45	47	1	10	18.8
6 LIFE	84(H)	21(L)	5	64	61	67	67	65	5	40	53	2	10	18.8
7 G40	84	24	6	62	60	64	63	64	6	44	48	2	12	18.8
8 72 Overall	100%	27%	14%	59	57	62	62	61	5%	38%	55%	3%	13%	18.8

Note. — (H) High; (L) Low.

[a]*See description in Table B5.*

Table B5. **Enlistment Standards Description and Abbreviation**

Standard Description	Abbreviation
1. Current Air Force Enlistment Standards require a minimum combined total of 170 on the four aptitude composites (Mechanical, Administrative, General, and Electrical) of the Armed Service Vocational Aptitude Battery (ASVAB)	G45/170
2. Minimum combined total of 165 on the four aptitude composites of the ASVAB; minimum score of 40 on the General Aptitude composite; minimum age of 18 years.	G40/165/18
3. Minimum Combined total of 150 on the four aptitude composites of the ASVAB; minimum score 40 on the General Aptitude composite.	G40/150
4. Minimum combined total of 165 on the four aptitude composites of the ASVAB; minimum age of 18 years.	165/18
5. Minimum combined total of 165 on the four aptitude composites of the ASVAB.	165
6. Standard derived by weighting the factors described earlier in the paper by the appropriate coefficients and using a cut off score of .52.	LIFE
7. Minimum score of 40 on the General Aptitude composite.	G40
8. Actual standard used for 1972 accession. Minimum score of 40 on at least two of the four aptitude composites of the ASVAB.	72 Overall

Note. – All standards except LIFE assume that if an applicant is classified as Mental Category III or IV on the Armed Forces Qualifying Test he/she must be a high school graduate.

All standards except 72 Overall simulate the current Category IV restriction of one per recruiting detachment per month (i.e., approximately 40 per month nationwide.

☆U.S. GOVERNMENT PRINTING OFFICE:1979– 671-056/21